We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

2

Application of Principal Component Analysis to Elucidate Experimental and Theoretical Information

Cuauhtémoc Araujo-Andrade et al.*
Unidad Académica de Física, Universidad Autónoma de Zacatecas
México

1. Introduction

Principal Component Analysis has been widely used in different scientific areas and for different purposes. The versatility and potentialities of this unsupervised method for data analysis, allowed the scientific community to explore its applications in different fields. Even when the principles of PCA are the same in what algorithms and fundamentals concerns, the strategies employed to elucidate information from a specific data set (experimental and/or theoretical), mainly depend on the expertise and needs of each researcher.

In this chapter, we will describe how PCA has been used in three different theoretical and experimental applications, to explain the relevant information of the data sets. These applications provide a broad overview about the versatility of PCA in data analysis and interpretation. Our main goal is to give an outline about the capabilities and strengths of PCA to elucidate specific information. The examples reported include the analysis of matured distilled beverages, the determination of heavy metals attached to bacterial surfaces and interpretation of quantum chemical calculations. They were chosen as representative examples of the application of three different approaches for data analysis: the influence of data pre-treatments in the scores and loadings values, the use of specific optical, chemical and/or physical properties to qualitatively discriminate samples, and the use of spatial orientations to group conformers correlating structures and relative energies. This reason fully justifies their selection as case studies. This chapter also pretends to be a reference for those researchers that, not being in the field, may use these methodologies to take the maximum advantage from their experimental results.

* Claudio Frausto-Reyes[2], Esteban Gerbino[3], Pablo Mobili[3], Elizabeth Tymczyszyn[3], Edgar L. Esparza-Ibarra[1], Rumen Ivanov-Tsonchev[1] and Andrea Gómez-Zavaglia[3]
[1]*Unidad Académica de Física, Universidad Autónoma de Zacatecas*
[2]*Centro de Investigaciones en Óptica, A.C. Unidad Aguascalientes*
[3]*Centro de Investigación y Desarrollo en Criotecnología de Alimentos (CIDCA)*
[1,2]*México*
[3]*Argentina*

2. Principal component analysis of spectral data applied in the evaluation of the authenticity of matured distilled beverages

The production of distilled alcoholic beverages can be summarised into at least three steps: *i)* obtaining and processing the raw materials, *ii)* fermentation and distillation processes, and *iii)* maturation of the distillate to produce the final aged product (Reazin, 1981). During the obtaining and fermentation steps, no major changes in the chemical composition are observed. However, throughout the maturation process, distillate undergoes definite and intended changes in aromatic and taste characteristics.

These changes are caused by three major types of reactions continually occurring in the barrel: *1)* extraction of complex wood substances by liquid (*i.e.:* acids, phenols, aldehydes, furfural, among others), *2)* oxidation of the original organic substances and of the extracted wood material, and *3)* reaction between various organic substances present in the liquid to form new products (Baldwin et al., 1967; Cramptom & Tolman,1908; Liebman & Bernice, 1949; Rodriguez-Madera et al., 2003; Valaer & Frazier,1936). Because of these reactions occurring during the maturation process, the stimulation and odour of ethanol in the distillate are reduced, and consequently, its taste becomes suitable for alcoholic beverages (Nishimura & Matsuyama, 1989). It is known that the concentration of extracts from wood casks in matured beverages seriously depend on the casks conditions (Nose et al., 2004). Even if their aging periods are the same, the use of different casks for the maturation process, strongly conditions the concentration of these extracts. (Philip, 1989; Puech, 1981; Reazin, 1981). Diverse studies on the maturation of distillates like whiskey, have demonstrated that colour, acids, esters, furfural, solids and tannins increase during the aging process. Except for esters, the greatest rate of change in the concentration of these compounds occurs during the first year (Reazin, 1981). For this reason, the extracts of wood and the chemically produced compounds during the aging process confer some optical properties that can be used to evaluate the authenticity and quality of the distillate in terms of its maturation process (Gaigalas et al., 2001; Walker, 1987).

The detection of economic fraud due to product substitution and adulteration, as well as health risk, requires an accurate quality control. This control includes the determination of changes in the process parameters, adulterations in any ingredient or in the whole product, and assessment that flavours attain well defined standards. Many of these quality control issues have traditionally been assessed by experts, who were able to determine the quality by observing their colour, texture, taste, aroma, etc. However, the acquisition of these skills requires years of experience, and besides that, the analysis may be subjective. Therefore, the use of more objective tools to evaluate maturation becomes essential. Nevertheless, it is difficult to find direct sensors for quality parameters. For this reason, it is necessary to determine indirect parameters that, taken individually, may weakly correlate to the properties of interest, but as a whole give a more representative picture of these properties. In this regard, different chromatographic techniques provide reliable and precise information about the presence of volatile compounds and the concentration of others (*i.e.:* ethanol, methanol, superior alcohols or heavy metals, etc.), thus proving the quality and authenticity of distilled alcoholic beverages (Aguilar-Cisneros, et al., 2002; Bauer-Christoph et al., 2003; Ragazzo et al.,2001; Savchuk et al., 2001; Pekka et al., 1999; Vallejo-Cordoba et al., 2004). In spite of that, chromatographic techniques, generally destroy the sample under study and also require equipment installed under specific protocols and installations

(Abbott & Andrews, 1970). On the other hand, the use of spectroscopic techniques such as infrared (NIR and FTIR), Raman, ultraviolet/visible together with multivariate methods, has already been used for the quantification of the different components of distilled beverages (*i.e.:* ethanol, methanol, sugar, among others). This approach allows the evaluation of quality and authenticity of these alcoholic products in a non-invasive, easy, fast, portable and reliable way (Dobrinas et al., 2009; Nagarajan et al., 2006). However, up to our knowledge, none of these reports has been focused on the evaluation of the quality and authenticity of distilled beverages in terms of their maturation process.

Mezcal is a Mexican distilled alcoholic beverage produced from agave plants from certain regions in Mexico (NOM-070-SCFI-1994), holding origin denomination. As many other similar matured distilled beverages, mezcal can be adulterated in the flavour and appearance (colour), these adulterations aiming to imitate the sensorial and visual characteristics of the authentic matured beverage (Wiley, 1919). Considering that the maturation process in distillate beverages has a strong impact on their taste and price, adulteration of mezcal beverage pursuit obtaining the product in less time. However, the product is of lower quality. In our group, a methodology based in the use of UV-absorption and fluorescence spectroscopy has been proposed for the evaluation of the authenticity of matured distilled beverages, and focused in mezcal. We took advantage of the absorbance/emission properties of woods extracts and molecules added to the distilled during maturation in the wood casks. In this context, principal component analysis method appears as a suitable option to analyse spectral data aiming to elucidate chemical information, thus allowing discrimination of authentic matured beverages from those non-matured or artificially matured.

In this section, we present the PCA results obtained from the investigation of two sets of spectroscopic data (UV absorption and fluorescence spectra), collected from authentic mezcal samples at different stages of the maturation: *white or young* (non-maturated), *rested* (matured ≥ 2 months in wood casks), and *aged* (≥1 year in wood casks). Samples belonging to false matured mezcals (artificially matured) are labelled as: *abocado* (white or young mezcal artificially coloured and flavoured) and *distilled* (coloured white mezcal). These samples were included with the aim of discriminating authentic matured mezcals from those artificially matured. The discussion is focused on the influence of the pre-treatments of spectra on the scores and loadings values. The criteria used for the scores and loadings interpretation are also discussed.

2.1 Spectra pre-treatment

Prior to PCA, spectra were smoothed. Additionally, both spectra data sets were mean centred (MC) prior the analysis as a default procedure. In order to evaluate the effect of the standardization pre-treatment (1/Std) over the scores and loadings values, PCA was also conducted over the standardized spectra. Multivariate spectra analysis and data pre-treatment were carried out using The Unscrambler ® software version 9.8 from CAMO company.

2.2 Collection of UV absorption spectra

Spectra were collected in the 285-450 nm spectral range, using an UV/Vis spectrometer model USB4000 from the Ocean Optics company, coupled to the Deuterium tungsten

halogen light source and cuvette holder by means of optical fibers, and with a spectral resolution of ~1.5 nm. The mezcal samples were deposited in disposable 3.0 mL cuvettes, specially designed for UV/Vis spectroscopy under a transmission configuration, which remained constant for all measurements.

2.2.1 PCA-scores

Fig. 1 (a) and (b), depict the distribution of objects (samples/spectra) corresponding to the two pre-treatment options (MC and 1/Std) in the PC-space. In both graphs, a similar distribution of objects along PC1-axis was observed. The groupings along PC1 indicate a good discrimination between matured and non-matured mezcals. Additionally, samples corresponding to mezcals *a priori* known as artificially matured (*i.e.* abocado and distilled samples) and other few, labeled as rested but presumably artificially matured, cluster together with the non-maturated ones. This indicates that the UV absorbance properties of compounds and molecules naturally generated in the wood cask, are significantly different from those from other compounds used with counterfeit purposes (Boscolo et al, 2002).

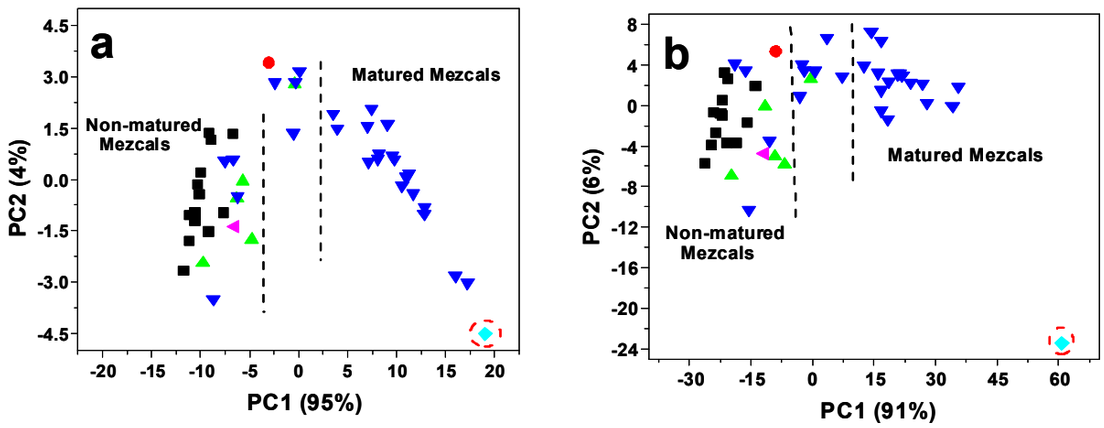

Fig. 1. PCA-Scores plots obtained from raw UV absorption spectra (a) mean centred, (b) standardized. (■) White/young, (●) White w/worm, (▲) abocado or artificially matured, (◄) distilled (white/young coloured), (▼) rested and (♦) aged.

A central region, delimited with dashed lines and mainly including samples corresponding to rested mezcals and a few artificially mature samples (abocado and white mezcal w/worm) can be considered as an "indecisive zone". However, taking into account that some samples analysed in this study were directly purchased from liquor stores, it may be possible that few of them, claimed as authentic rested, have been artificially matured. In addition, the sample corresponding to aged mezcal is separated from all the other samples in both graphs, but always clustering together with the rested samples. This indicates that the cluster of objects/samples is related not only with their maturation stage, but also with their maturation time. This behaviour points out that the standardization pre-treatment does not affect significantly the distribution of objects in the scores plots. However, there are some issues that must be considered: in Fig. 1 (a), the aged sample is located close to the rested group, but non as part of it. This can be explained in terms of their different times of maturation. On the other hand, in Fig. 1 (b), the aged sample seems to be an outlier or a

sample non-related with the other objects. This unexpected observation can be explained considering the similarity between the spectra of the rested and aged mezcals [see Fig. 2 (a)]. For this reason, the PCA-scores plot corresponding to standardized spectra, must be considered cautiously since they can lead to incorrect interpretations.

2.2.2 PCA-loadings

Once the distribution of objects in the PC-space has been interpreted, the analysis of the one-dimensional loadings plots has been carried out in order to find the relationship between the original variables (wavelength) and the scores plots (Esbensen, 2005; Geladi & Kowalski, 1986; Martens & Naes, 1989). In this case, PC1 is the component discriminating mezcal samples according to their maturation stage and time. Consequently, the PC1-loadings provide information about the spectral variables contributing to that discrimination. Fig.2 (a) shows four representative absorption spectra in the 290-450 nm ultraviolet range for white, abocado, rested and aged mezcals. According to the Figure, the absorption spectra of white and abocado samples look similar, and different from those corresponding to the rested and aged mezcals.

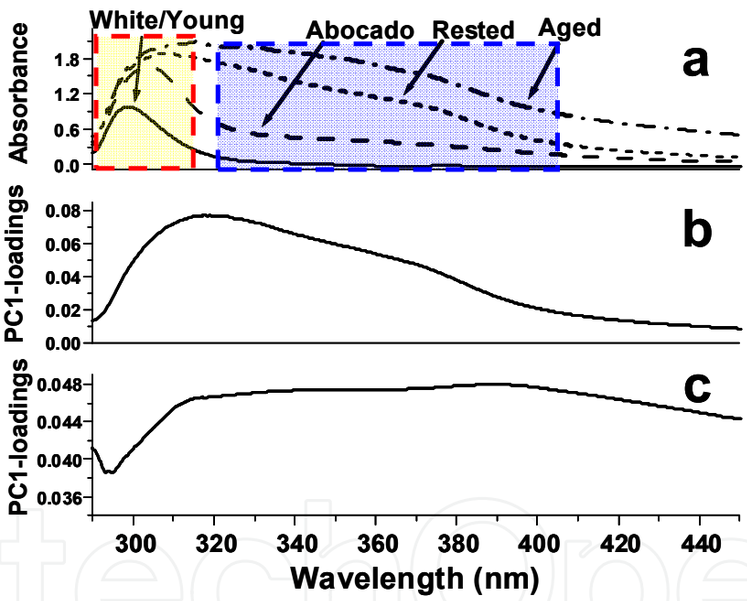

Fig. 2. (a) Representative raw UV absorption spectra for each of the four types of mezcals, (b) PC1-loadings plot for the centred spectra, and (c) PC1-loadings plot for the centred and standardized spectra.

The loading plots indicate that the 320-400 nm region [blue dashed rectangle, Fig. 2 (a)], is the best region to evaluate the authenticity of matured mezcals because the wood compounds extracted, produced and added to mezcals during the aging process absorb in this region. The 290-320 nm range [red dashed rectangle, Fig. 2 (a)], provides the signature for non-maturated mezcals. Fig. 2 (b) and (c) depict one-dimensional PC1 loadings plots corresponding to mean centred and standardized spectra, respectively. From Fig. 2 (b), it is feasible to observe the great similarity between the one-dimensional PC1 loadings plot and the representative spectrum of rested and aged mezcals, suggesting that PC1 mainly models

the spectral features belonging to authentic matured mezcals. On the other hand, one-dimensional loadings plot obtained from standardized spectra [Fig. 2 (c)], lacks in the spectral information provided, thus limiting its uses for interpretation purposes. In spite of that, standardization may be useful for certain applications (*i.e.* calibration of prediction/classification by PLS or PLS-DA) (Esbensen, 2005).

2.3 Collection of fluorescence spectra

Taking into account that the emission spectra of organic compounds can provide information about them and about their concentration in mixed liquids, this spectroscopic technique appears as a complementary tool allowing the evaluation of the authenticity of matured alcoholic beverages (Gaigalas et al., 2001; Martínez et al, 2007; Navas & Jimenez, 1999; Walker, 1987). Fluorescence spectra were collected in the 540-800 nm spectral range, using a spectrofluorometer model USB4000-FL from the Ocean Optics company, coupled to a laser of 514 nm wavelength and cuvette holder by optical fibers. The spectral resolution was ~10 nm. The mezcal samples were put into 3.0 mL quartz cuvettes, in a 90 degrees configuration between the excitation source and the detector. This orientation remained constant during the collection of all the spectra. The laser power on the samples was 45 mW.

2.3.1 PCA-scores

Fig. 3 (a) and (b) depict the scores plots obtained from the mean centred spectra. According to Fig. 3 (a), PC1 explains 90 % of the variance. Two groups can be observed along PC1-axis, one of them including white mezcals and ethanol, and the other one, including rested, abocado and distilled mezcals. This indicates that data structure is mainly influenced by the presence or absence of certain organic molecules (not necessarily extracted from wood), all of them having similar emission features.

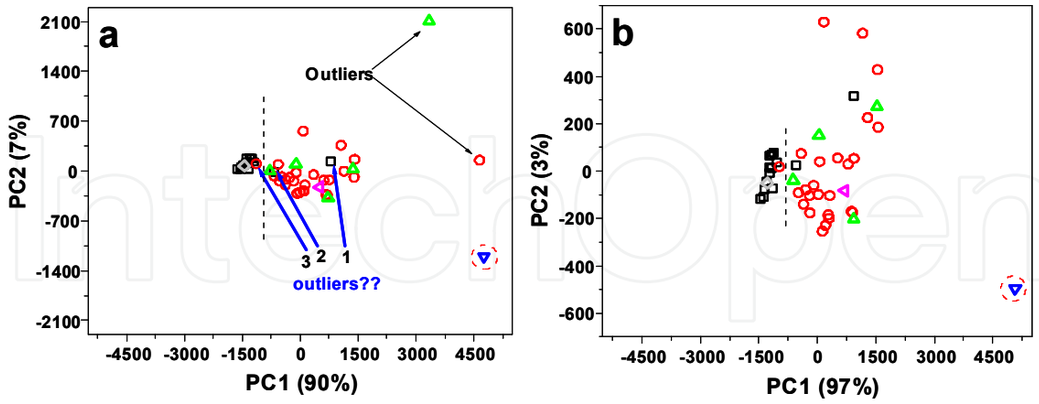

Fig. 3. PCA-scores plots obtained for the mean centred fluorescence spectra. Samples correspond to different stages of maturation. (a) Scores plot before the removal of outliers, (b) scores plot after the removal of outliers. (□) White or young, (△) abocado, (○) rested, (▽) aged, (◁) distilled and (◇) ethanol.

Three isolated objects, corresponding to rested, abocado and aged, can also be observed along PC1-axis. Among them, the first two can be considered as outliers. On the contrary, in

the case of the aged sample, the higher concentration of wood extracts in comparison to the rested samples originates a noticeably different spectrum, thus explaining the observation of this aged sample as an isolated object. In order to improve the distribution of objects, samples detected as outliers were removed. Fig. 3 (b) shows the scores plot after outliers removal. Similarly to Fig. 3(a), two main groups can be observed. However, the percentage of explained variance for PC1 increases to 97%. This fact indicates that, even when the object distribution does not depict significant variations, the removal of these outliers, allows the model to describe the data structure in a more efficient way.

There are other samples that can be considered as outliers [indicated with numbers 1-3 in Fig. 3(a)]. Among them, number 1 corresponds to white mezcal with worm (some mezcal producers add a worm to their product as a distinctive), the worm probably providing certain organic molecules. These organic molecules would have similar emission properties than those of the rested ones. Hence, number 1 can be taken as outlier or not. We decide to take it, as correctly classified. On the contrary, objects 2 and 3, were considered as outliers, because they do not have any particular characteristics like object 1. Furthermore, when they were removed from the PCA (data not shown), distribution of objects and explained variance percentages remained similar. For this reason, we decided not to remove them.

In conclusion, only the joint analysis of the scores plot, the raw spectra, the loading plots and all other available information of the samples can give a correct interpretation of our results.

Figure 4 (a) and (b) shows the PCA-scores plots obtained from standardized spectra before and after the removal of outliers. As it was described before, the similar object distribution observed in Fig. 3 and 4 indicates that standardization does not provide any additional benefit for the spectral data analysis.

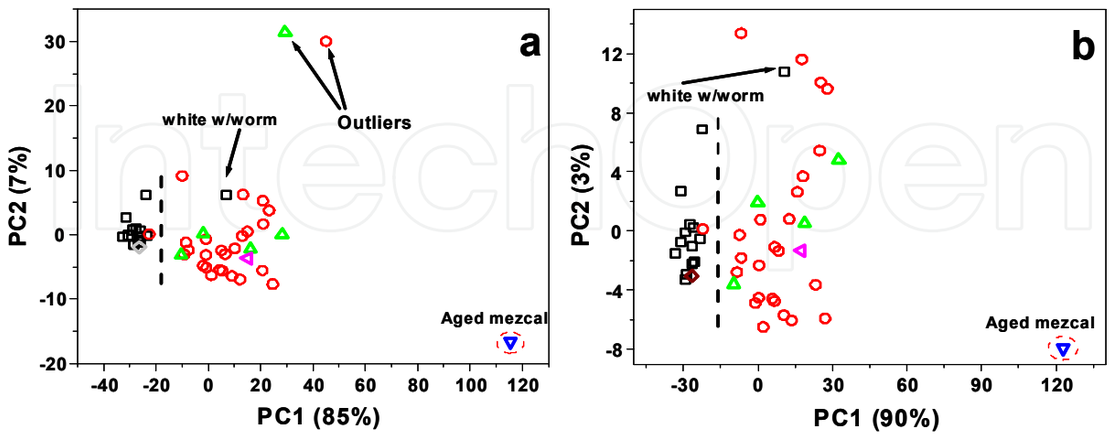

Fig. 4. PCA-scores plots obtained for the mean centred and standardized fluorescence spectra. Samples correspond to different stages of maturation. (a) Scores plot before the removal of outliers, (b) scores plot after the removal of outliers. (□) White or young, (△) abocado, (○) rested, (▽) aged, (◁) distilled and (◇) ethanol.

2.3.2 PCA-loadings

Fig. 5 (a) depicts six representative fluorescence spectra corresponding to each type of mezcal analysed. A high similarity between the fluorescence spectrum of ethanol and white/young mezcal is observed. On the other hand, rested, abocado and distilled mezcals, have similar spectra. Finally, the huge differences between the intensity of the emission spectra corresponding to aged mezcal, and that of the other types of mezcal, can be attributed to the higher concentration of organic molecules coming from the wood cask during the maturation process.

This also explains the grouping along PC1-axis. From these results, it can be concluded that PC1 can discriminate between naturally/artificially matured samples and white mezcal samples. On the contrary, the one-dimensional PC1-loading plot obtained from standardized spectra; does not provide clear information about the objects distribution.

In this sense, and according with the results described above, the standardization pre-treatment does not improve the discrimination between samples in the scores plots. On the contrary, it leads to a misinterpretation of the loading plots.

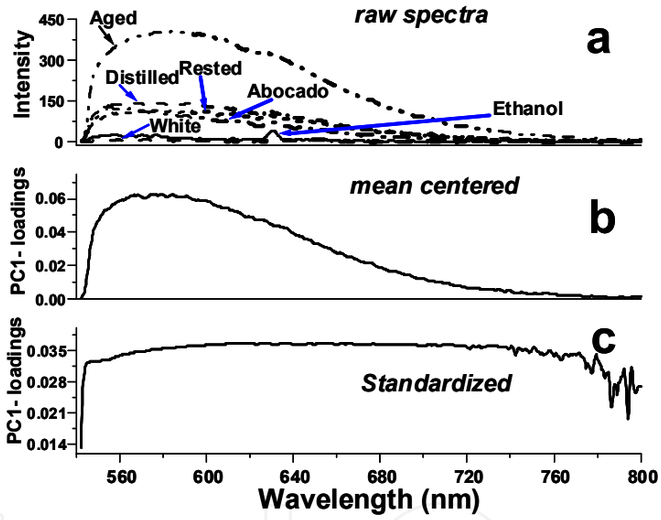

Fig. 5. (a) Representative raw fluorescence spectra for each type of mezcals, (b) PC1-loadings plot for the centred spectra, and (c) PC1-loadings plot for standardized spectra.

2.4 Final remarks

The results described above, showed that PCA conducted over a set of UV absorption spectra from different types of mezcals, allows an efficient and reliable discrimination between artificially and naturally maturated mezcals in wood casks, the data pre-treatments playing an important role for the correct interpretation. This discrimination power is based on the differential absorbance spectra of the compounds naturally produced from wood during maturation stage and those corresponding to the compounds used for adulteration (*i.e.:* colorants to confer a matured appearance to the beverage).

On the other hand, PCA conducted over fluorescence spectra allowed the identification of two main groups correlated with the presence or absence of certain organic molecules, not

necessarily correlated with maturation in the wood casks. Thus, fluorescence spectroscopy did not demonstrate to have enough sensibility to discriminate between authentic and artificially matured mezcals.

3. PCA of Raman spectra for the determination of heavy metals attached to bacterial surfaces

Toxic metals are not degradable and tend to accumulate in the exposed organisms causing serious health effects. The use of biological agents to remove or neutralize contaminants (biorremediation) is a very important tool for the removal of such toxics. In particular, the use of inactivated microorganisms as adsorbents (biosorption) has been suggested as an effective and economical way to remove heavy metals from water and food intended to human or animal consumption (Davis et al., 2003; Haltunen et al., 2003, 2007, 2008; Ibrahim et al. 2006; Mehta & Gaur, 2005; Mrvčić et al., 2009; Shut et al., 2011; Volesky & Holan, 1995).

Metal biosorption is usually evaluated by means of analytical methods or atomic absorption spectrometry. These methods allow the quantification of free metal ions in the supernatants of bacterial/metal samples (Ernst et al., 2000; Haltunen et al., 2003, 2007, 2008; Velazquez et al., 2009; Zolotov et al., 1987).

In this sense, a method involving vibrational spectroscopic techniques (*i.e.:* Raman spectroscopy) and multivariate methods (both unsupervised and/or supervised), would represent an advantage over the standard procedures, due to the possibility of quantifying the metal ions directly from the bacterial sample, and at the same time, obtaining structural information (Araujo-Andrade et al., 2004, 2005, 2009; Ferraro et al., 2003; Gerbino et al. 2011; Jimenez Sandoval, 2000).

PCA carried out on the Raman spectra represents the first step in the construction of a calibration model allowing the quantification of metal ions attached to bacterial surfaces. This analysis allows obtaining a correlation between the spectral variations and the property of interest (*i.e.* the metal ion concentration), identifying the optimal spectral region/s for the calibration of quantification models, and also detecting erroneous measurements leading to reduce the predictive ability of the model.

In this section, we present the PCA results obtained from the Raman spectra corresponding to bacterial samples (*Lactobacillus kefir*) before and after the interaction with four heavy metals (Cd^{2+}, Pb^{2+}, Zn^{2+}, Ni^{2+}) in three different concentrations each.

Even when the main objective of this study was to calibrate models for the quantification of metal ions attached to bacterial surfaces using supervised methods (*i.e.*: PLS), the calibration of prediction models goes beyond of the intention of this chapter. For this reason, we focused this section just on the discussion of the PCA results.

3.1 Collection of spectral data set

The Raman spectra of the bacterial samples before and after the interaction with metal ions were measured by placing them onto an aluminum substrate and then under a Leica microscope (DMLM) integrated to a Renishaw micro-Raman system model 1000B. In order

to retain the most important spectral information from each sample, multiple scans were conducted in different points of the bacterial sample moving the substrate on an X-Y stage.

The Raman system was calibrated with a silicon semiconductor using the Raman peak at 520 cm^{-1}, and further improved using samples of chloroform ($CHCl_3$) and cyclohexane (C_6H_{12}). The wavelength of excitation was 830 nm and the laser beam was focused on the surface of the sample with a 50X objective.

The laser power irradiation over the samples was 45 mW. Each spectrum was registered with an exposure of 30 seconds, two accumulations, and collected in the 1800-200 cm^{-1} region with a spectral resolution of 2 cm^{-1}.

3.2 Spectral data pre-treatment

Raman spectra analyzed were collected over dry solid bacterial samples before and after the interaction with different concentrations of metal ions. Therefore, it is highly probable that our measurements include some light scattering effects (background scattering). These effects are in general composed of multiplicative and additive effects (Martens & Naes, 1989).

Spectra collected and analyzed in this section were baseline corrected in order to subtract the fluorescence contribution. To perform this correction, a polynomial function was approximated to the spectrum baseline, and after that, subtracted from the spectrum. Also, the spectra were smoothed using Savitzky-Golay method. Light scattering effects were corrected using the multiplicative scatter correction (MSC) algorithm and then, the spectra were mean centred. Data pre-treatment and multivariate spectra analysis were carried out with Origin version 6.0 from Microcal Company, and The Unscrambler® software version 9.8 from CAMO company.

3.3 Analysis and discussion of PCA results

PCA was performed on the pre-treated Raman spectra of each bacteria/metal sample in order to correlate metal concentrations with the spectral information.

The criteria used for PCA-scores and loadings interpretation are depicted in the next subsection. Even when the presented data set corresponds to the bacteria/Cd^{+2} interaction, the same methodology was employed for the analysis of the other bacteria/metal samples.

3.3.1 Scores interpretation

Fig. 6 depicts the PCA-scores plots obtained before and after outliers exclusion (panels a and b, respectively). Three main groups, labeled as I, II and III, can be observed in Fig. 6 (a). These groups can be represented by their PC-coordinates as follows: (+*i*, +*j*), (-*i*, -*j*) and (-*i*, +*j*), where *i* and *j* represent the *i-esime* and *j-esime* score value for PC1 and PC2, respectively.

These three groups or clusters are highly related with the three different concentrations of cadmium trapped on the bacterial surfaces. However, there are several potential outliers that should be removed to get a better description of the data structure. In this case, the outliers were selected on the basis of the dispersion existent among objects of the same group along PC1, the component explaining the major percentage of variance of the data set.

After the removal of outliers, the three groups identified in Fig. 6 (a), became much better defined in the PC-space. However, a different cluster distribution in the PC-space was observed [Fig. 6 (b)]. Table 1 describes these changes in terms of their PC- coordinates before and after the removal of outliers.

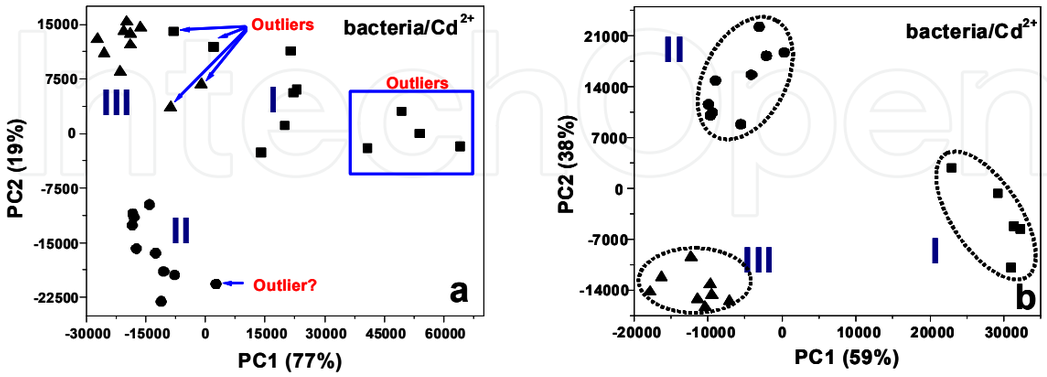

Fig. 6. PCA-Scores plots obtained from pre-treated Raman spectra corresponding to three concentrations of bacteria/Cd^{+2} samples: (■) 0.059 mM, (●) 0.133 mM, (▲) 0.172 mM. (a) With outliers, (b) without outliers.

Additionally, a different distribution of the individual percentage of explained variances was observed in both PCs (PC1, 77% to 59%, and PC2 19% to 38%). However, the total percentage of explained variances before and after the removal of outliers was similar (96%before and 97% after the removal of outliers). This indicates that the removal of outliers did not reduce the information about the data structure provided by both PCs.

	$(PC1_i, PC2_j)$ coordinates before outlier removal	$(PC1_i, PC2_j)$ coordinates after outlier removal
Cluster I	(+*i*, +*j*)	(+*i*, -*j*)
Cluster II	(-*i*, -*j*)	(-*i*, +*j*)
Cluster III	(-*i*,+*j*)	(-*i*, -*j*)

Table 1. Cluster coordinates in the PC-space before and after the removal of outliers.

According to Fig 6 (b), a good discrimination between the lowest (group I) and the medium/highest cadmium concentrations (groups II and III) was observed along PC1-axis.

In summary, it can be concluded that PC1 allows a gross discrimination (due to the huge difference in the concentration of samples clustered in I and samples clusters in II and III). Lower differences in Cd^{+2} concentrations are modelled by PC2 (clusters II and III are well separated along this PC).

3.3.2 Loadings interpretation

Once that distribution of objects in the scores plot was interpreted and correlated with the cadmium concentration attached to the bacterial biomass, the one-dimensional loadings

before and after the removal of outliers were analysed to correlate Cd^{+2} concentrations with changes in the original spectra (*i.e.* Raman shift, wavenumber, wavelength, etc.).

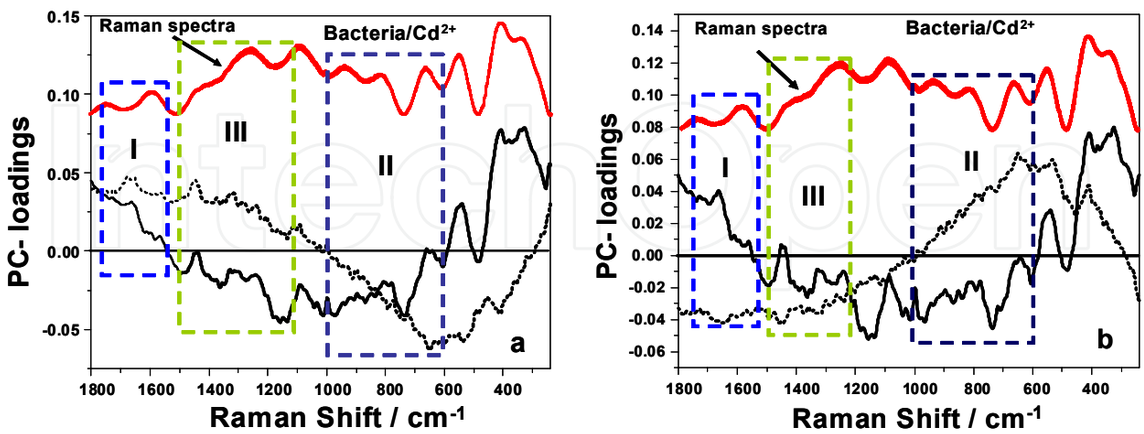

Fig. 7. One-dimensional loadings plots obtained from the PCA corresponding to different bacteria/Cd^{+2} concentrations; (a) before and (b) after the removal of outliers. Solid line corresponds to PC1-loadings and dashed line to PC2-loadings.

Considering that PCs can be represented as a linear combination of the original unit vectors, where the loadings are the coefficients in these linear combinations, distribution and/or localization of each object in the PC-space has a direct relation with their respective PC-loadings values (Esbensen, 2005).

Fig. 7 (a) depicts the one-dimensional loadings plots corresponding to PC1 and PC2 before the removal of outliers. The influent spectral regions for the distribution of objects in the PC-space were underlined using dashed frames. The main spectral differences between objects of cluster I and objects of clusters II and III were found in the 1800-1500 cm^{-1} spectral region. This region was selected taking into account the loadings values and the PC-coordinates for each cluster.

For instance, cluster I has PC-coordinates (+*i*, +*j*), then we selected the region where both loadings, PC1 and PC2 have positive values (in this case, the 1800-1500 cm^{-1} region). For cluster II, whose coordinates are -*i*, -*j*, we selected the region with negative loadings values for both PCs (1000-600 cm^{-1}). Finally, for cluster III, we selected the region with negative and positive loadings values for PC1 and the PC2, respectively (1500-1100 cm^{-1}) whose coordinates are -*i*, +*j*. The same strategy was adopted for the loading analysis after the removal of outliers [Figure 7 (b)]. Even when the loading values are different, the spectral regions representing each cluster are the same.

In summary, it can be concluded that the main spectral differences between objects of cluster I and objects of clusters II and III, can be observed in the 1800-1500 cm^{-1} region. Very interestingly, the carboxylate (COO^{-}) groups absorb in this region. In our previous work we have reported that metal ions can be attached to the bacterial surface through the COO^{-} groups (Gerbino et al., 2011). Spectral differences between clusters II and III were found in the 1000-600 cm^{-1} and 1500-1100 cm^{-1} regions, respectively.

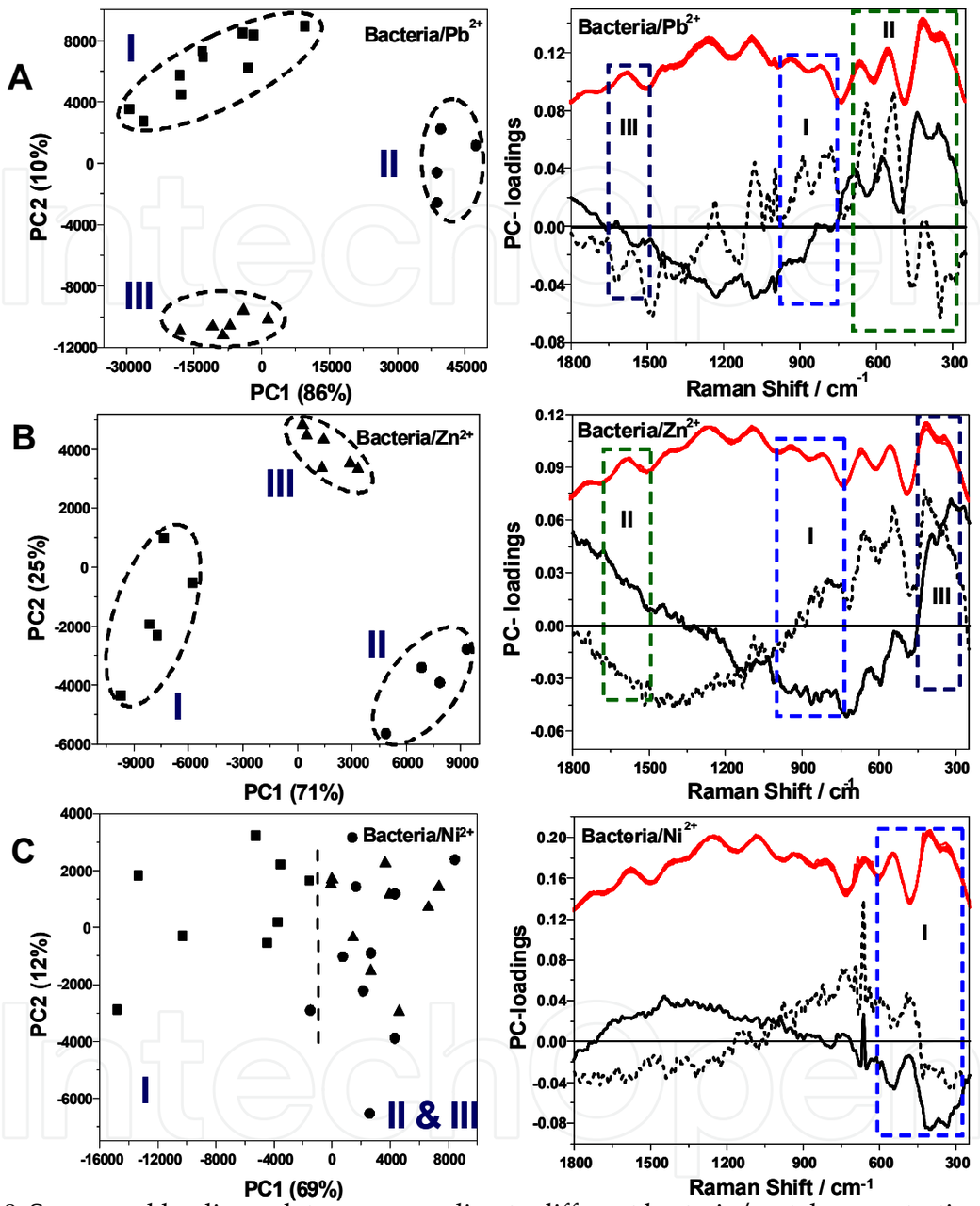

Fig. 8. Scores and loadings plots corresponding to different bacteria/metal concentrations. a) lead: (■) 0.028 mM, (●) 0.181 mM, (▲) 0.217 mM, b) zinc: (■) 0.114 mM, (●) 0.307 mM, (▲) 0.350 mM and c) nickel: (■) 0.022 mM, (●) 0.109 mM, (▲) 0.181 mM. Solid line corresponds to PC1-loadings and dashed line to PC2-loadings.

These two regions provide vibrational information about the phosphate groups and superficial polysaccharides related with certain bacteria superficial structures that have been previously reported as responsible for the bacteria/metal interaction (Mobili et al., 2010; Sara & Sleytr, 2000).

3.3.3 PCA results for other metal ions

The same approach performed for bacteria/Cd^{+2} interaction was employed to analyse the interaction of the other metal ions with the bacterial surface. The scores and loadings plots obtained after the removal of outliers from PCA, carried out on the spectral data sets corresponding to samples of different concentrations of the three metal ions attached to the bacterial surfaces are shown in Fig. 8. A clear discrimination between objects corresponding to different concentrations of lead and zinc trapped on the bacterial surface was observed along the PC-space [Figure 8(a) & (b)]. The information provided by the loadings explains the distribution of samples in clusters surrounded by ellipses and labeled as I, II and III.

The dashed rectangles depicted in the loadings plots indicate the spectral regions that are statistically influent in the distribution of objects. For lead, region labeled as I corresponds to the superficial polysaccharides (1000-800 cm^{-1} region), region labeled as II, to the fingerprint region (700-250 cm^{-1}), and region labeled as III, to the amide I region (1650-1500 cm^{-1}). For zinc, region I corresponds to superficial polysaccharides (1000-750 cm^{-1}), region II, to amide I (1700-1500 cm^{-1}), and region III, to the fingerprint region (450-250 cm^{-1}).

It is important to point out that in both bacteria/Pb^{+2} and bacteria/Zn^{+2} interactions, the same spectral regions were identified as influent for the objects distribution observed in PCA. This indicates that similar molecular structures may be involved in the bacteria/metal interaction. The scores plot corresponding to the bacteria/Ni^{+2} samples indicate a poor discrimination among objects [Fig. 8 (c)]. According to the loadings plots of the bacteria/Ni^{+2} samples, the spectral similarities and differences between samples are mainly in the fingerprint region (region surrounded by a blue dashed line).

3.4 Final remarks

From the results discussed and described above, it can be concluded that Raman spectra allow obtaining chemical information related with bacteria/metal interactions, and also with metal ion concentrations. PCA allowed an efficient elucidation of the information obtained from the spectra. Furthermore, the one-dimensional loadings analysis allowed identifying the influent spectral regions, as well as the molecular structures involved in the objects/samples distribution in the scores plots.

4. PCA applied to the interpretation of quantum chemical calculations

Quantum chemical calculations in large flexible molecules represent a challenge, due to the high number of combinations of conformationally relevant parameters and computational resources/capabilities required. This is an extremely complicated task, unless a systematic approach is carried out. In this section, a PCA-based methodology allowing the correlation between molecular structures and properties of a conformationally flexible molecule (arbutin) is described. This procedure is simple and requires relatively modest computational facilities (Araujo-Andrade et al. 2010).

Arbutin is an abundant solute in the leaves of many freezing- or desiccation-tolerant plants. It has been used pharmaceutically in humans for centuries, either as plant extracts or, in more recent decades, in the purified form. Arbutin acts as an antiseptic or antibacterial agent on the urinary mucous membranes while converting into hydroquinone in the kidney

(Witting et al., 2001). It is also used as a depigmenting agent (skin whitening agent) as it inhibits melanin synthesis by inhibition of tyrosinase activity.

From a chemical point of view, arbutin is a flexible molecule composed by a glucopyranoside moiety bound to a phenol ring (Fig. 9). It has eight conformationally relevant dihedral angles, five of them related with the orientation of the hydroxyl groups and the remaining three taking part in the skeletal of the molecule.

Up to our knowledge, no attempts to use of a PCA based methodology for the structural analysis of quantum chemical information were reported.

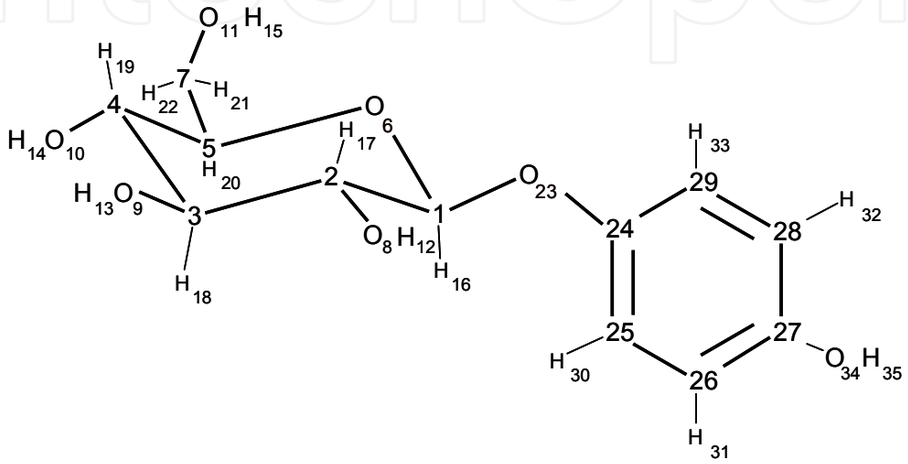

Fig. 9. Arbutin molecule, with atom numbering scheme. (copyrighted from Araujo-Andrade et al., 2010)

4.1 Quantum chemical calculations

The semi-empirical PM3 method (Stewart, 1989) was used to perform a systematic preliminary conformational search on the arbutin potential energies surface (PES), which were later on taken into account in the subsequent, more reliable analysis performed at higher level of theory. This preliminary conformational search was carried out using the HyperChem Conformational Search module (Howard & Kollman, 1988; HyperChem, Inc. © 2002; Saunders, 1987, 1990).

The eight dihedral angles defining the conformational isomers of arbutin (Fig. 9) were considered in the random search: $C_2C_1O_{23}C_{24}$, $C_1O_{23}C_{24}C_{25}$, $O_6C_5C_7O_{11}$, $C_5C_7O_{11}H_{15}$, $C_3C_4O_{10}H_{14}$, $C_2C_3O_9H_{13}$, $C_1C_2O_8H_{12}$ and $C_{26}C_{27}O_{34}H_{35}$. Conformations with energies lower than 50 kJ mol^{-1} were stored while higher-energy conformations or duplicate structures were discarded. The structures obtained from this conformational search were used as start points for the construction of the input files later used in the higher level quantum chemical calculations. These latter were performed with Gaussian 03 (Gaussian, 2003) at the DFT level of theory, using the 6-311++G(d,p) basis set (Frisch et al, 1990) and the three-parameter density hybrid functional abbreviated as B3LYP, which includes Becke's gradient exchange correction (Becke, 1988) and the Lee, Yang and Parr (Lee et al, 1988) and Vosko, Wilk and Nusair correlation functionals (Vosko et al., 1980). Conformations were optimized using the Geometry Direct Inversion of the Invariant Subspace (GDIIS) method (Csaszar & Pulay,

1984). The optimized structures of all conformers were confirmed to correspond to true minimum energy conformations on the PES by inspection of the corresponding Hessian matrix. Vibrational frequencies were calculated at the same level of theory. PCA were performed using The Unscrambler® software (v9.8).

4.2 Theoretical data set and pre-treatment

The group of Cartesian coordinates corresponding to the 35 atoms of arbutin (see Fig. 9), for each of the 130 conformers found after the conformational analysis was used as data set in this study. In other words, our data set consisted of a matrix of 130 x 105 elements, corresponding to the arbutin conformers and the x, y, z coordinates of each atom of the molecule, respectively. Data were mean centred prior PCA.

4.3 Data analysis

In order to provide a general and fast procedure to perform the PCA on the conformational data sets, the next strategy was followed: *1)* In the Cartesian referential, all conformers were oriented, in such a way that the structurally rigid fragment of arbutin (the glucopyranoside ring) was placed as close as possible to the axes origin; *2)* All Cartesian coordinates of the 130 conformers of arbutin were then used to perform the PCA. The table of data (data matrix) was built as follows: each row corresponds to a conformer and the columns to the Cartesian coordinates: the first 35 columns, to the x- coordinates of the 35 atoms of arbutin, the second 35 columns, to the y coordinates, and the last 35 columns, to the z coordinates.

4.3.1 Scores and loadings analysis

Fig. 10 depicts the distribution of the arbutin conformers, in the PC-space. Three well separated groups were observed. Group A (squares), appears well separated from groups B (circles) and C (triangles) along the PC1-axis, which explain 77% of total variance. Group B is separated from Group C along the PC2-axis. In order to elucidate the main structural parameters for this separation, the one-dimensional loadings values were analyzed. The loading values of PC1 and PC2, indicate the atoms' coordinates contributing the most to structurally distinguish the conformers of arbutin in the chosen reference system [Fig. 10 (b) and (c), respectively].

According to these values, the orientation of the atoms 25 to 35, related with the spatial orientation of the phenol ring relatively to the reference glucopyranoside fragment, is the main contributing factor allowing for the differentiation among conformers. Consequently, the relative spatial orientation of the phenol ring is determined by the dihedrals interconnecting the glucopyranoside and phenol rings, $C_2C_1O_{23}C_{24}$ and $C_1O_{23}C_{24}C_{25}$, which are then shown to be of first importance in structural terms. Fig. 10 (d) shows the means and standard errors of the means (standard deviation of the sample divided by the square root of the sample size) of the 8 conformationally relevant dihedral angles of arbutin (the angles were first converted to the 0-360^o range). From this graph, it can be clearly observed that $C_2C_1O_{23}C_{24}$ and $C_1O_{23}C_{24}C_{25}$ dihedral angles, describe the distribution of the three groups along the PC1 and PC2 axis, respectively. In other words, PC1 is related with the $C_2C_1O_{23}C_{24}$ dihedral angle, and allows the discrimination of conformers belonging to group A. In these conformers, the phenol ring is placed above and nearly perpendicular to the

glucopyranoside ring. On the contrary, in all conformers belonging to groups B and C, the phenol ring is pointing out of the glucopyranoside moiety, and oriented to the side of the oxygen atom from the glucopyranoside ring. PC2 allows a specific discrimination among the three groups of conformers. This specificity factor is given by the values of $C_1O_{23}C_{24}C_{25}$.

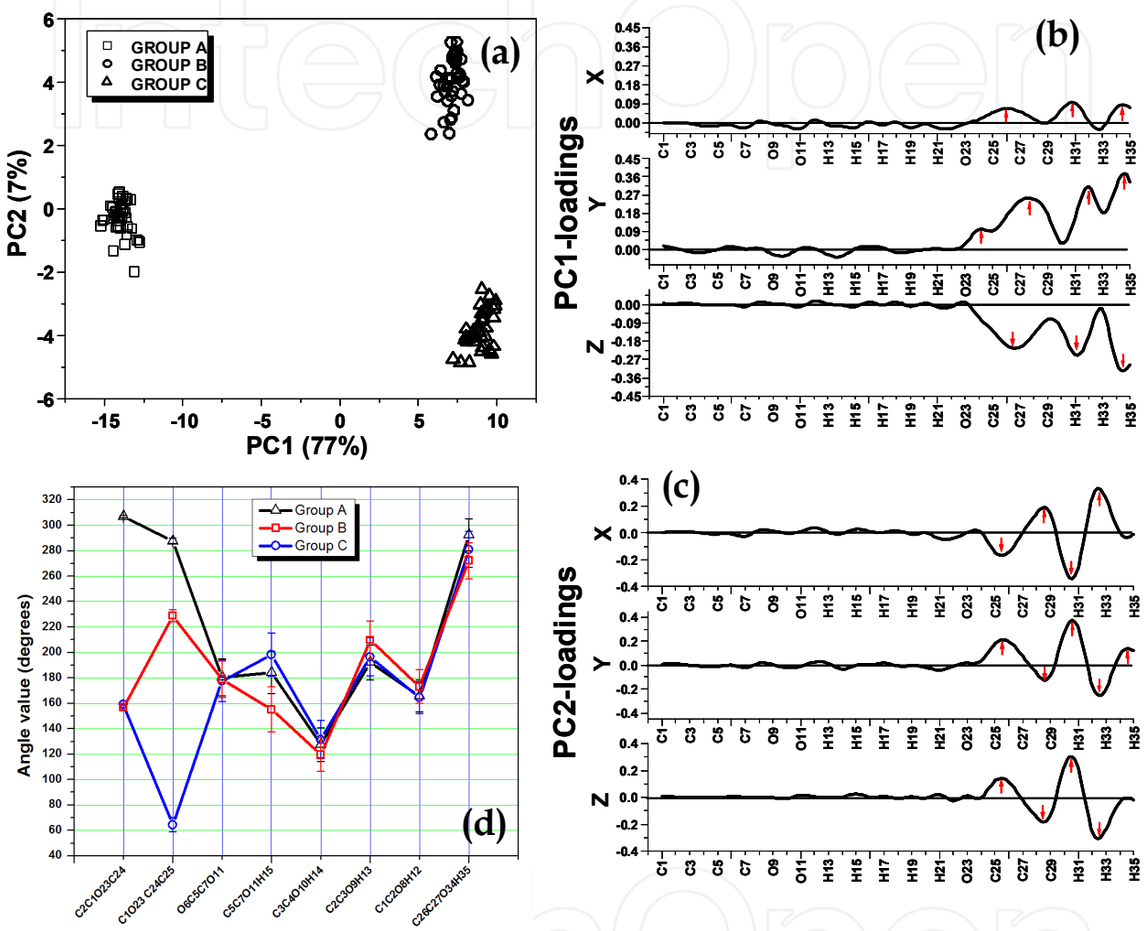

Fig. 10. (a) PCA-scores and, (b, c) the corresponding loadings grouping arbutin conformers in terms of structural similarity. (d) total average values and standard deviations of the 8 conformationally relevant dihedral angles of arbutin in the 3 groups of conformers. (copyrighted from Araujo-Andrade et al., 2010)

The relationship between the energetic and conformational parameters related with each of the three groups identified in the scores plot, was also investigated. Fig. 11 depicts the relative energy values (taken as reference the energy of the conformational ground state) for each conformer according to the group they belong. From an energetic point of view, groups B and C are equivalent. However, no conformer with relative energy below 15 kJ mol^{-1} belonging to Group A. This trend can be correlated with the orientations adopted by the phenol ring relatively to the glucopyranoside ring, as was described before.

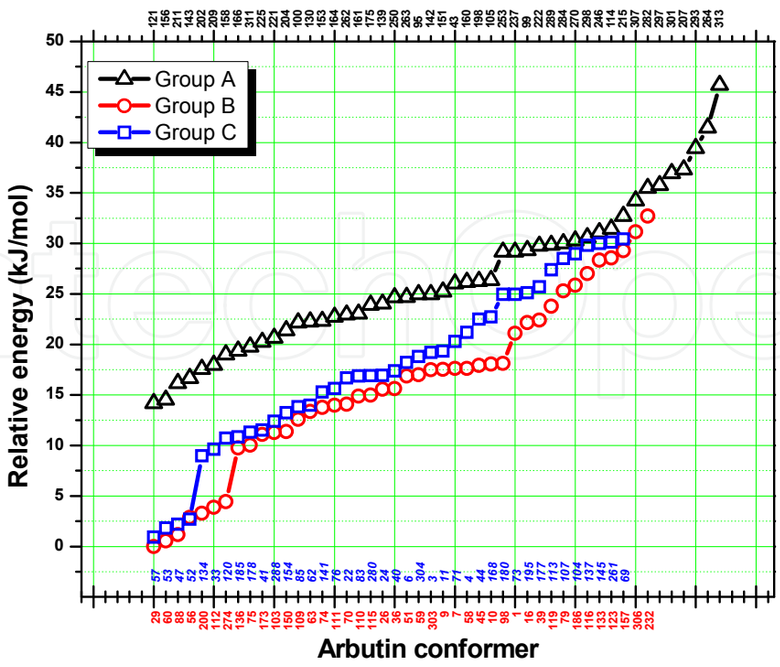

Fig. 11. Relative energies of the 130 lowest energy conformers of arbutin (the energy of the conformational ground state was taken as reference). (copyrighted from Araujo-Andrade et al., 2010)

Once the influence of the relative position of the two rings in arbutin on the relative energy of the conformers was evaluated, the preferred conformations assumed by the substituents of the glucopyranoside ring and their influence on energies were investigated in deeper detail. To this aim, PCA was conducted on each of the previously determined groups of conformers (A, B, C), excluding the x, y, z, coordinates corresponding to the phenol ring (atoms 23-35). This strategy allowed for the elimination of information that is not relevant for a conformational analysis within the glucopyranoside ring. PCA-scores/loadings analysis and interpretation was realized by using the methodology described above for the whole arbutin molecule. The results of this analysis are shown in Fig. 12-15. The PCA scores plot for Group A [Fig. 12 (a)] shows four well defined groups, labeled as subgroups A1, A2.a, A2.b and A3. If all elements of these four subgroups are projected over the PC1 axis, three groups can be distinguished, with one of them constituted by subgroups A1 and A3, other formed by subgroup A2.a and the third one by subgroup A2.b. On the other hand, projecting the elements over the PC2-axis allows also to distinguish three groups, but this time corresponding to A1, (A2.a, A2.b) and A3.

The observation of the one-dimensional loadings plots for PC1 and PC2 [Fig. 12 (b) & (c)], allows us to conclude that the positions of atoms C_7, H_{21}, H_{22}, O_{11} and H_{15} are highly related with the conformers distribution in the PCA scores plot, *i.e.*, the conformation exhibited by the CH_2OH substituent at C_5 is the main discriminating factor among subgroups. In consonance with this observation, when the mean values of the conformationally relevant dihedral angles associated with the substituted glucopyranoside ring in each subgroup are plotted [Fig. 12 (d)], it is possible to observe that the dihedral angles associated with the

CH_2OH substituent (O_6-C_5-C_7-O_{11} and C_5-C_7-O_{11}-H_{15}) are those allowing the discrimination among the four subgroups. In the case of the C_5-C_7-O_{11}-H_{15} dihedral, one can promptly correlate the three groups corresponding to the projection of the PCA subgroups over the PC1-axis with the three dihedral mean values shown in the plot: *ca*. 60, 160 and 275º (-85º), respectively for A2.b, A2.a and (A1, A3).

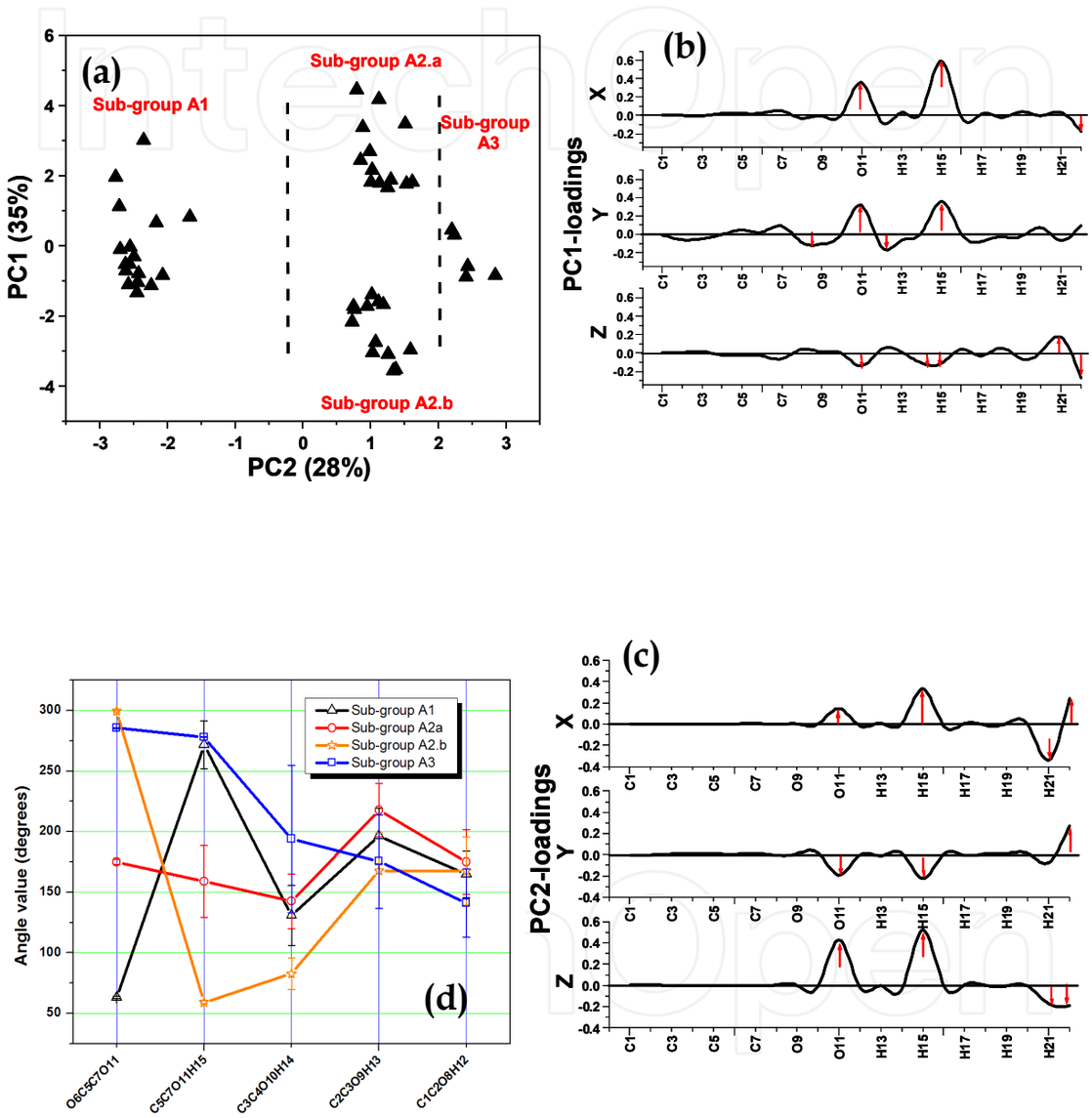

Fig. 12. (a) PCA-scores and, (b,c) the corresponding loadings and belonging to Group A in terms of structural similarity in the conformations of the substituents of the glucopyranoside ring. (d) total average values and standard errors of the means of the 5 conformationally relevant dihedral angles of the glucopyranoside ring arbutin in the 4 subgroups. (copyrighted from Araujo-Andrade et al., 2010)

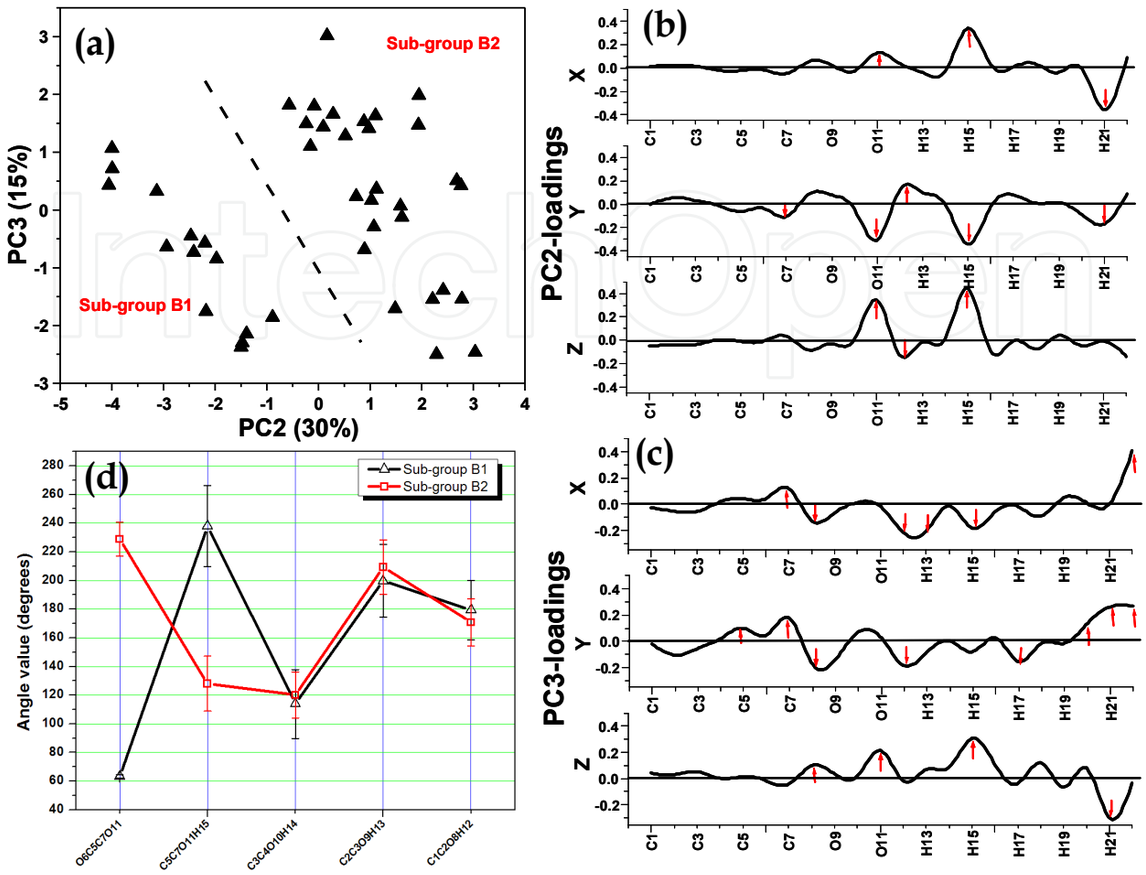

Fig. 13. (a) PCA-scores and, (b, c) the corresponding loadings belonging to Group B in terms of structural similarity in the conformations of the substituents of the glucopyranoside ring. (d) total average values and standard errors of the means of the 5 conformationally relevant dihedral angles of the glucopyranoside ring arbutin in the 2 subgroups of conformers. (copyrighted from Araujo-Andrade et al., 2010)

The PCA-scores plot for the conformers belonging to the Group B [Fig. 13 (a)] shows only two clear groupings of conformers, where PC2 is the component separating these two groups the best. The loadings plot of PC2 [Fig. 13 (b)] shows that the clusters are also determined by the positions of atoms C_7, H_{21}, H_{22}, O_{11} and H_{15}, *i.e.*, by the conformation of the CH_2OH fragment. As expected, these observations are in agreement with the dihedral angles' mean values plot [Fig. 13 (d)], which clearly reveals that there, the values of the O_6-C_5-C_7-O_{11} and C_5-C_7-O_{11}-H_{15} dihedral angles are the ones that mainly discriminate internal coordinates among the conformers belonging to subgroups B1or B2.

A similar analysis made for conformers belonging to Group C allows concluding that three subgroups (C1, C2 and C3) can be defined [Fig. 14 (a)], once again resulting mainly from different conformations assumed by the CH_2OH substituent [Fig. 14 (b) & (d)]. Regarding the energies of the conformers, subgroups are not strongly discriminative. However, subgroups A3, C2 and, in less extent B1, include conformers gradually less stable than the remaining subgroups of each main group (data not shown).

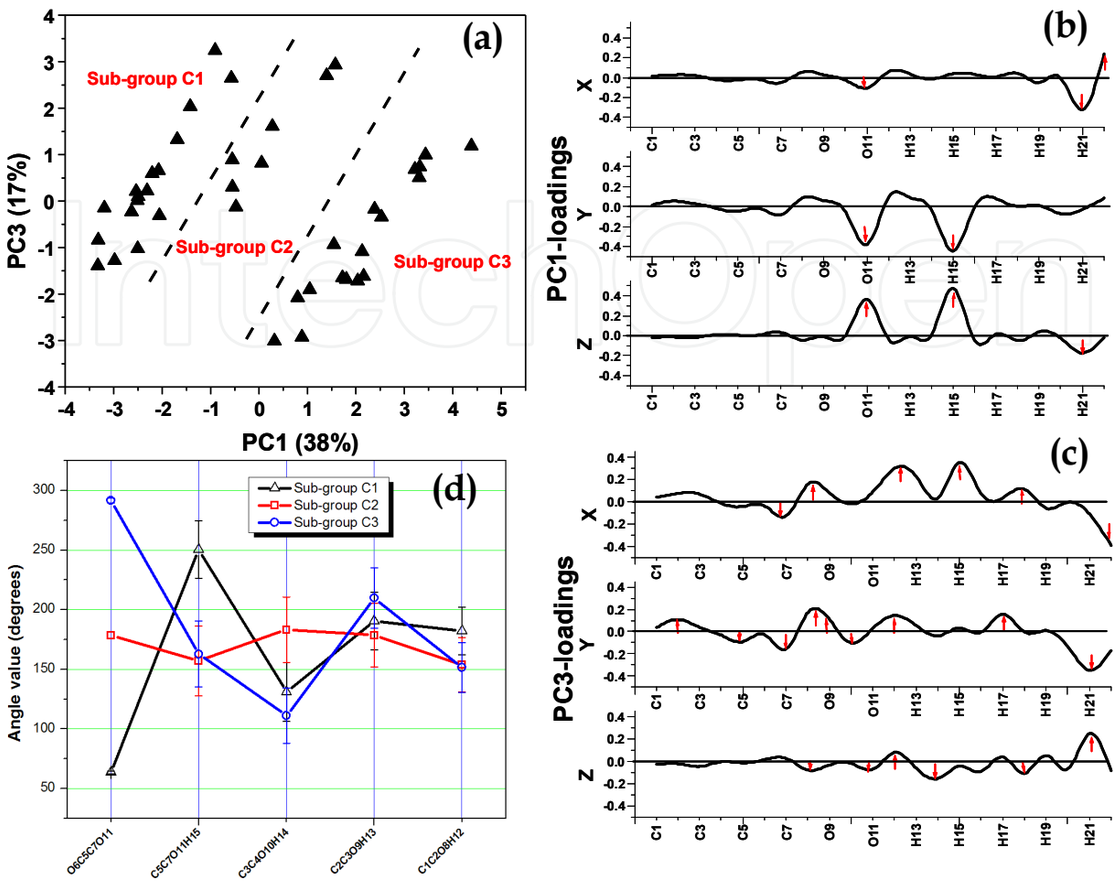

Fig. 14. (a) PCA-scores, and (b, c) the corresponding loadings, belonging to Group C. (d) total average values and standard errors of the means of the 5 relevant dihedral angles of the glucopyranoside ring arbutin in the 3 subgroups. (copyrighted from Araujo-Andrade et al., 2010)

4.4 Final remarks

PCA analyses based on atomic Cartesian coordinates of the properly oriented in the Cartesian system conformers of arbutin allowed the grouping of these conformers by structural analogies, which could be related with the conformationally relevant dihedral angles. Among them, the dihedrals interconnecting the glucopyranoside and phenol rings and those associated with the CH_2OH fragment were found to be the most relevant ones.

In summary, this work represents a new simple approach for the structural analysis of complex molecules and its aim was also to show another application of PCA.

5. Conclusion

The results reported in this chapter for each experimental and theoretical application of PCA, demonstrate the versatility and capabilities of this unsupervised method to analyse samples from different origins. Three different examples were selected to show the

relevance of PCA to elucidate specific information from a data collection in several fields. Among these issues, the following aspects must be underlined: a) the influence of the data pre-treatment on the scores and loadings values; b) the *a-priori* knowledge of the data source to select the appropriate data pre-processing; c) the strategies and criteria used for the scores and loadings plots interpretation and, d) criteria used for outliers detection, and their influence in the PCA model.

The amalgamation of the different sections included in this chapter can be used as a starting point for those researchers who are not specialists in the field, but that are interested in using these methodologies, to take the maximum advantage from their results.

6. Acknowledgments

This work was supported by CONACyT, Mexico [Projects No. 119491 (2009) and No. 153066 (2010)] and PROMEP, Mexico (Project UAZ-PTC-092), Agencia Nacional de Promoción Científica y Tecnológica, Argentina (Projects PICT/2008/145 and PICT/2010/2145), CYTED Program (Ciencia y Tecnología para el Desarrollo) Network P108RT0362 and CONACyT-CONICET (México, Argentina) (bilateral project res. Nº 962/07-05-2009), PIFI, México (project P/PIFI 2010-32MSU0017H-06). AGZ, PM and EET are members of the research career CONICET (National Research Council, Argentina). EG is doctoral fellow from CONICET.

7. References

Abbott, D. & Andrews, R. S. (1970). *Introduction to Chromatography*, Longman group LTD, ISBN 978-0582321946, London.

Aguilar-Cisneros, B. O.; López, M. G.; Richling, E.; Heckel, F.; Schreier, P. (2002). Tequila authenticity assessment by headspace SPME-HRGC-IRMS analysis of 13C/12C and 18O/16O ratios of ethanol. *J. Agric. Food Chem.*, Vol. 50, No.6, pp.7520-7523.

Araujo-Andrade, C., Ruiz, F., Martinez-Mendoza, J.R., and Terrones, H. (2004). Non-invasive in-vivo blood glucose levels prediction using near infrared spectroscopy. *AIP Conf. Proc.* Vol.724, pp. 234-239.

Araujo-Andrade, C., Campos-Cantón, I., Martínez, J.R., Ortega, G., and Ruiz, F. (2005). Prediction model based on multivariate analysis to determine concentration of sugar in solution. *Rev. Mex. Fis. E,* Vol.51, pp. 67-73.

Araujo-Andrade, C., Ruiz, F., Martínez-Mendoza, J.R., Padrón, F., and Hernández-Sierra, J. (2009). Feasibility for non invasive estimation of glucose concentration in newborns using NIR spectroscopy and PLS. *Trends Appl. Spectrosc.* Vol.7, pp. 27-37.

Araujo Andrade, C.; Lopes, S.; Fausto, R. and Gómez-Zavaglia, A. (2010). Conformational study of arbutin by quantum chemical calculations and multivariate analysis. *Journal of Molecular Structure*. Vol.975, pp. 100-109.

Baldwin, S. R., Black, A., Andreasen, A. A. & Adams, S. L. (1967). Aromatic congener formation in maturation of alcoholic distillates, *J. Agr. Food Chem.,* Vol.15, No.3, pp. 381-385.

Bauer-Christoph, C.; Christoph, N.; Aguilar-Cisneros, B. O.; López, M. G.; Richling, E.; Rossmann, A.; Schreier, P. (2003). Authentication of tequila by gas chromatography and stable isotope ratio analyses. *Eur. Food Res. Technol*., Vol.217, No.5, pp. 438-443.

Becke, A. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A*, Vol.38, pp. 3098-3100.

Boscolo, M.; Andrade-Sobrinho, L. G.; Lima-Neto B. S.; Franco D. W. & Castro Ferreira M. M. (2002). Spectrophotometric Determination of Caramel Content in Spirits Aged in Oak Casks, *Journal of AOAC International* Vol.85, No.3, pp. 744-750.

Crampton, C. A. & Tolman, L. M. (1908). A study of the changes taking place in Whyskey stored in wood, *J. Am. Chem. Soc.* Vol.30, No.1, pp. 98–136.

Csaszar, P. & Pulay, P. (1984). Geometry optimization by direct inversion in the iterative subspace, *J. Mol. Struct. (Theochem),* Vol.114, pp. 31-34.

Davis, T.A., Volesky, B., & Mucci, A. (2003). A review of the biochemistry of heavy metal biosorption by brown algae. *Water Research* Vol.37, pp. 4311-4330.

Dobrinas, S.; Stanciu, G. & Soceanu, A. (2009). Analytical characterization of three distilled drinks, *Ovidius University Annals Chemistry*, Vol.20, No.1, pp. 48-52.

Ernst, T.; Popp, R. & Van Eldik, R. (2000). Quantification of heavy metals for the recycling of waste plastics from electrotechnical applications, *Talanta* Vol.53, No.2, pp. 347–357.

Esbensen, K. H. (2005). *Multivariate Data Analysis - In Practice*, CAMO Software AS, ISBN 82-993330-3-2, Esbjerg, Denmark.

Ferraro, J. R.; Nakamoto, K. & Brown, C. W. (2003). Introductory Raman Spectroscopy, Academic Press, ISBN 978-0122541056, London.

Frisch,M.; Head-Gordon, M. & Pople, J. (1990). Semidirect Algorithms for the Mp2 Energy and Gradient, *Chem. Phys. Lett*. Vol.166, pp. 281-289.

Gaigalas, A.; K., Li, L.; Henderson, O.; Vogt, R.; Barr, J.; Marti, G.; Weaver, J.; Schwartz, A. (2001). The development of fluorescence intensity standards, *J. Res. Natl. Inst. Stand. Technol.* Vol.106, No.2, pp. 381-389.

Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004. Copyright C 1994-2003, Gaussian, Inc.

Geladi, P. & Kowalski, B. R. (1986). Partial Least Square Regression: A tutorial*, Anal. Chim. Acta*, Vol.185, pp. 1-17.

Gerbino, E., Mobili, P., Tymczyszyn, E.E., Fausto, R., and Gómez-Zavaglia, A. (2011). FTIR spectroscopy structural analysis of the interaction between Lactobacillus kefir S-layers and metal ions. *Journal of Molecular Structure*. Vol.987, pp. 186-192.

Halttunen, T., Kankaanpää, P., Tahvonen, R., Salminen, S. & Ouwehand, A.C. (2003). Cadmium removal by lactic acid bacteria. *Bioscience Microflora* Vol. 22, pp.93-97.

Halttunen, T., Salminen, S. & Tahvonen, R. (2007). Rapid removal of lead and cadmium from water by specific lactic acid bacteria. *International Journal of Food Microbiology* Vol.114, pp. 30-35.

Halttunen, T., Salminen, S., Meriluoto, J., Tahvonen, R., & Lertola, K. (2008). Reversible surface binding of cadmium and lead by lactic acid & bifidobacteria. *International Journal of Food Microbiology* Vol.125, pp.170-175.

Hincha, D.; Oliver, A. E. & Crowe, J. H. (1999). Lipid composition determines the effects of arbutin on the stability of membranes, *Biophys. J.* Vol.77, No.4. pp.2024-2034.

Howard, A. E.; & Kollman, P. A. (1988). An analysis of current methodologies for conformational searching of complex molecules, *J. Med. Chem*. Vol.31, No.9, pp.1669-1675.

HyperChem Conformational Search module (2002). Tools for Molecular Modeling. Hypercube, Inc., 1115 NW 4th St., Gainesville, FL 32608 (USA)

Ibrahim, F., Halttunen, T., Tahvonen, R., & Salminen, S. (2006). Probiotic bacteria as potential detoxification tools: assessing their heavymetal binding isotherms. *Canadian Journal of Microbiology* Vol.52, pp. 877-885.

Jimenez Sandoval, S. (2000). Micro-Raman spectroscopy: a powerful technique for materials research, *Microelectronic Journal*, Vol.31, pp. 419-427.

Lee, C.; Yang, W. & Parr, R. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B.* Vol.37, pp. 785-789.

Liebmann, A. J. & Scherl, B. (1949). Changes in whisky while maturing, *Industrial and engineering chemistry*, Vol.41, No.3, pp. 534-543.

Martens, H. & Næs, T. (1989). *Multivariate Calibration*, Wiley & Sons, ISBN 0-471-90979-3, Chichester, England.

Martínez, J. R.; Campos-Cantón, I.; Martínez-Castañón, G.; Araujo-Andrade, C. & Ruiz, F. (2007). Feasibility of laser induced fluorescence as a rapid method for determination of the time stored of aged alcoholic beverages, *Trends in applied spectroscopy*, Vol.6, pp. 28-33.

Martínez, J. R.; Campos-Cantón, I.; Araujo-Andrade, C.; Martínez-Castañón, G. & Ruiz, F. (2007). Analysis of Mexican spirit drinks mezcal using near infrared spectroscopy, *Trends in Applied Spectroscopy*, Vol.6, pp. 35-41.

Mehta, S.K. & Gaur, J.P. (2005). Use of algae for removing heavy metal ions from wastewater: progress and prospects. *Critical Reviews in Biotechnology* Vol.25, pp.113-152.

Mobili, P.; Londero, A.; De Antoni, G.; Gomez-Zavaglia, A.; Araujo-Andrade, C.; Ávila-Donoso, H.; Ivanov-Tzonchev, R.; Moreno-Hernandez, I. & Frausto Reyes, C. (2010). Multivariate analysis of Raman spectra applied to microbiology: discrimination of microorganisms at the species level, *Revista Mexicana de Física*, Vol.56, No.5, pp. 378–385.

Mobili, P.; Araujo-Andrade, C.; Londero, A.; Frausto-Reyes, C.; Ivanov-Tzonchev, R.; De Antoni, G.L.; & Gómez-Zavaglia, A. (2011). Development of a method based on chemometric analysis of Raman spectra for the discrimination of heterofermentative lactobacilli. *Journal of Dairy Research* Vol.78, pp. 233-241.

Mrvčić, J., Prebeg, T., Barišić, L., Stanzer, D., Bačun-Družina, V. & Stehlik-Tomas, V. (2009). Zinc binding by lactic acid bacteria. *Food Technology Biotechnology* Vol.47, pp. 381-388.

Nagarajan, R.; Gupta, A.; Mehrotra, R.; & Bajaj, M. M. (2006). Quantitative Analysis of Alcohol, Sugar, and Tartaric Acid in Alcoholic Beverages Using Attenuated Total Reflectance Spectroscopy, Journal of Automated Methods and Management in Chemistry, *J Autom Methods Manag Chem.* Vol.2006, No. pp. 1–5.

Navas, M. J. & Jiménez, A. M. (1999). Chemoluminescent methods in alcoholic beverage analysis, *J. Agric. Food Chem*. Vol.47, No.1, pp. 183-189.

Nishimura, K. & Matsuyama R. (1989). Maturation and maturation chemistry, In: *the science and technology of Whiskies*, J.R. Piggott, R. Sharp, E. E. B. Duncan, (Ed.), 244-253, Longman Scientific & Technical, ISBN 978-0582044289, Essex, U.K.

Norma Oficial Mexicana NOM-070-SCFI-1994, Bebidas alcohólicas-Mezcal-Especificaciones.

Nose, A., Hojo, M., Suzuki, M. & Ueda, T. J. (2004). Solute effects on the interaction between water and etanol in aged whiskey, Agric. Food Chem. Vol.(52), pp. 5359-5365.

Pekka, J. L.; LaDena A. K. & Eero T. A-M. (1999). Multi-method analysis of matured distilled alcoholic beverages for brand identification, *Z Lebensm Unters Forsch*, Vol. 208, pp. 413-417.

Philp, J. M. (1989). Cask quality and warehouse condition, In: *The Science and Technology of Whiskies*; Piggott, J.R., Sharp, R., Duncan, R.E.B., (Ed.), 270-275, Longman Scientific & Technical, ISBN 978-0582044289, Essex, U.K.

Puech, J.L. (1981). Extraction and evolution of lignin products in Armagnac matured in oak, *Am. J. Enol. Vitic*. Vol.32, No.2, pp. 111-114.

Ragazzo, J. A.; Chalier, P.; Crouzet, J.; Ghommidh, C. (2001). Identification of alcoholic beverages by coupling gas chromatography and electronic nose. *Spec. Publ.-R. Soc. Chem.: Food FlaVors Chem*., Vol. 274, pp. 404-411.

Reazin, G. H. (1981). Chemical mechanisms of whiskey maturation, *Am. J. Enol. Vitic.* Vol.32, No.4, pp. 283-289.

Rodriguez Madera, R., Blanco Gomis, D. & Mangas Alonso, J. J. (2003). Influence of distillation system, oak wood type, and aging time on composition of cider brandy in phenolic and furanic compounds, *J. Agr. Food. Chem.* Vol.51, No.27, pp. 7969-7973.

Sara, M. & Sleytr, U.B. (2000). S-layer proteins. *Journal of Bacteriology* Vol.182, pp.859-868.

Saunders, M. (1987). Stochastic Exploration of Molecular Mechanics Energy Surfaces. Hunting for the Global Minimum, *J. Am. Chem. Soc*. Vol.109, No.10, pp.3150-3152.

Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, J. M.; Chang ,G. & Guidal, W. C. (1990). Conformations of cycloheptadecane. A comparison of methods for conformational searching, *J. Am. Chem. Soc.* Vol.112, No.4, pp. 1419-1427.

Savchuk, S. A.; Vlasov, V. N.; Appolonova, S. A.; Arbuzov, V. N.; Vedenin, A. N.; Mezinov, A. B.; Grigor'yan, B. R. (2001). Application of chromatography and spectrometry to the authentication of alcoholic beverages. *J. Anal. Chem*. Vol.56, No.3, pp. 214- 231.

Schut, S., Zauner, S., Hampel, G., König, H., & Claus, H. (2011). Biosorption of copper by wine-relevant lactobacilli. *International Journal of Food Microbiology* Vol.145, pp. 126-131.

Stewart, J. J. P. (1989). Optimization of Parameters for Semi-Empirical Methods I-Method, *J. Comput. Chem*. Vol.10, pp. 209-220.

Turner, S. R.; Senaratna, T.; Touchell, D. H.; Bunn, E.; Dixon, K. W. & Tan, B. (2001b). Stereochemical arrangement of hydroxyl groups in sugar and polyalcohol

molecules as an important factor in effective cyropreservation. *Plant Science* Vol.160, pp.489-497.

Valaer, P. & Frazier, W. H. (1936). Changes in whisky stored for four years, *Industrial and Engineering chemistry*, Vol.28, No.1, pp. 92-105.

Vallejo-Cordoba, B.; González-Córdova, A. F.; Estrada- Montoya, M. del C., (2004). Tequila volatile characterization and ethyl ester determination by solid-phase microextraction gas chromatography/ mass spectrometry analysis. *J. Agric. Food Chem*., Vol.52, No.18, pp. 5567-5571.

Velazquez, L. & Dussan, J. (2009). Biosorption and bioaccumulation of heavy metals on dead and living biomass of Bacillus sphaericus. *Journal of Hazardous Material*. Vol.167, pp.713-716.

Volesky, B. & Holan, Z.R. (1995). Biosorption of heavy metals. *Biotechnology Progress* Vol.11, pp.235-250.

Vosko, S.; Wilk, L. & Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, *Can. J. Phys.* Vol.58, No.8, pp. 1200-1211.

Walker, D. A. (1987). A fluorescence technique for measurements of concentration in mixing liquids, *J. Phys. E: Sci. Instrum*. Vol.20, No.2, pp. 217-24.

Wiley, H. W. (1919). *Beverages and their adulteration*, Campbell Press. ISBN 978-1443755740.

Wittig, J.; Wittemer, S. & Veit, M. (2001). Validated method for the determination of hydroquinone in human urine by high-performance liquid chromatography-coulometric-array detection*, J. Chromatogra. B*, Vol.761, No.1, pp. 125-132.

Zolotov, Y. A.; Malofeeva, G. I.; Petrukhin, O. M. & Timerbaev, A. R. (1987). New methods for preconcentration and determination of heavy metals in natural, *waterPure & Appl. Chem*., Vol.59, No.4, pp. 497-504.

Principal Component Analysis

Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0195-6
Hard cover, 300 pages
Publisher InTech
Published online 02, March, 2012
Published in print edition March, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as image processing, biometric, face recognition and speech processing. It also includes the core concepts and the state-of-the-art methods in data analysis and feature extraction.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Cuauhtémoc Araujo-Andrade Claudio Frausto-Reyes, Esteban Gerbino, Pablo Mobili, Elizabeth Tymczyszyn Edgar L. Esparza-Ibarra, Rumen Ivanov-Tsonchev and Andrea Gómez-Zavaglia (2012). Application of Principal Component Analysis to Elucidate Experimental and Theoretical Information, Principal Component Analysis, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0195-6, InTech, Available from: http://www.intechopen.com/books/principal-component-analysis/application-of-principal-component-analysis-to-elucidate-experimental-and-theoretical-information

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen